TRhizo-urbanMicrobiome

Pathogenic Bacteria

David Murray-Stoker

Contents

Load Packages

Load Data

```r
## Microbiome tidyamplicons
microbiome.tidyamplicon.reference <- read_rds(
    file = "data/microbiome_tidyamplicon_reference.rds"
)

## Load the urbanization data
urbanization.data <- read_rds(
    file = "data/urbanization_data.rds"
)

## Load the workspace
#load("data_analysis/10-bacterial_functional_groups/pathogenic_bacteria-workspace.RData")
```

Data Management

```r
## Set reference ASV abundance matrix
ASV.abundance.matrix <- add_rel_abundance(microbiome.tidyamplicon.reference) %>%
  abundances()

## Set reference taxa matrix
ASV.taxa.matrix <- taxa(microbiome.tidyamplicon.reference)

## Set tibble of bacteria abundance
bacteria.abundance.data <- ASV.abundance.matrix %>%
  full_join(ASV.taxa.matrix, by = "taxon_id") %>%
  full_join(microbiome.tidyamplicon.reference$samples, by = "sample_id") %>%
  select(Population:Compartment, phylum:genus, abundance, rel_abundance) %>%
  rename(
    Phylum = phylum, Class = class, Order = order, Family = family, Genus = genus,
    Abundance = abundance, Relative_Abundance = rel_abundance
  )

## Set aggregated sample data
sample.data <- microbiome.tidyamplicon.reference$samples %>%
    select(Population:Compartment) %>%
    distinct()
```

Filter Abundances

```r
## Set tibbles of abundances for different mutualistic bacteria
# Mycobacterium
mycobacterium.abundances <- bacteria.abundance.data %>%
  filter(Genus == "Mycobacterium") %>%
  select(Population, Compartment, Abundance, Relative_Abundance) %>%
  group_by(Population, Compartment) %>%
  summarise(
    Mycobacterium_Abundance = sum(Abundance),
    Mycobacterium_Relative_Abundance = sum(Relative_Abundance),
    .groups = "keep"
  )

# Legionella
legionella.abundances <- bacteria.abundance.data %>%
  filter(Genus == "Legionella") %>%
  select(Population, Compartment, Abundance, Relative_Abundance) %>%
  group_by(Population, Compartment) %>%
  summarise(
    Legionella_Abundance = sum(Abundance),
    Legionella_Relative_Abundance = sum(Relative_Abundance),
    .groups = "keep"
  )

# Staphylococcus
staphylococcus.abundances <- bacteria.abundance.data %>%
  filter(Genus == "Staphylococcus") %>%
  select(Population, Compartment, Abundance, Relative_Abundance) %>%
  group_by(Population, Compartment) %>%
  summarise(
    Staphylococcus_Abundance = sum(Abundance),
    Staphylococcus_Relative_Abundance = sum(Relative_Abundance),
    .groups = "keep"
  )

# Stenotrophomonas
stenotrophomonas.abundances <- bacteria.abundance.data %>%
  filter(Genus == "Stenotrophomonas") %>%
  select(Population, Compartment, Abundance, Relative_Abundance) %>%
  group_by(Population, Compartment) %>%
  summarise(
    Stenotrophomonas_Abundance = sum(Abundance),
    Stenotrophomonas_Relative_Abundance = sum(Relative_Abundance),
    .groups = "keep"
  )

# Acinetobacter
acinetobacter.abundances <- bacteria.abundance.data %>%
  filter(Genus == "Acinetobacter") %>%
  select(Population, Compartment, Abundance, Relative_Abundance) %>%
  group_by(Population, Compartment) %>%
  summarise(
```

```r
    Acinetobacter_Abundance = sum(Abundance),
    Acinetobacter_Relative_Abundance = sum(Relative_Abundance),
    .groups = "keep"
  )


# Clostridium
clostridium.abundances <- bacteria.abundance.data %>%
  filter(Genus == "Clostridium_sensu_stricto") %>%
  select(Population, Compartment, Abundance, Relative_Abundance) %>%
  group_by(Population, Compartment) %>%
  summarise(
    Clostridium_Abundance = sum(Abundance),
    Clostridium_Relative_Abundance = sum(Relative_Abundance),
    .groups = "keep"
  )


# Paenibacillus
paenibacillus.abundances <- bacteria.abundance.data %>%
  filter(Genus == "Paenibacillus") %>%
  select(Population, Compartment, Abundance, Relative_Abundance) %>%
  group_by(Population, Compartment) %>%
  summarise(
    Paenibacillus_Abundance = sum(Abundance),
    Paenibacillus_Relative_Abundance = sum(Relative_Abundance),
    .groups = "keep"
  )


# Williamsia
williamsia.abundances <- bacteria.abundance.data %>%
  filter(Genus == "Williamsia") %>%
  select(Population, Compartment, Abundance, Relative_Abundance) %>%
  group_by(Population, Compartment) %>%
  summarise(
    Williamsia_Abundance = sum(Abundance),
    Williamsia_Relative_Abundance = sum(Relative_Abundance),
    .groups = "keep"
  )


# Sphingomonas
sphingomonas.abundances <- bacteria.abundance.data %>%
  filter(Genus == "Sphingomonas") %>%
  select(Population, Compartment, Abundance, Relative_Abundance) %>%
  group_by(Population, Compartment) %>%
  summarise(
    Sphingomonas_Abundance = sum(Abundance),
    Sphingomonas_Relative_Abundance = sum(Relative_Abundance),
    .groups = "keep"
  )


# Xanthomonas
xanthomonas.abundances <- bacteria.abundance.data %>%
  filter(Genus == "Xanthomonas") %>%
  select(Population, Compartment, Abundance, Relative_Abundance) %>%
```

```r
  group_by(Population, Compartment) %>%
  summarise(
    Xanthomonas_Abundance = sum(Abundance),
    Xanthomonas_Relative_Abundance = sum(Relative_Abundance),
    .groups = "keep"
  )

# Rhizobacter
rhizobacter.abundances <- bacteria.abundance.data %>%
  filter(Genus == "Rhizobacter") %>%
  select(Population, Compartment, Abundance, Relative_Abundance) %>%
  group_by(Population, Compartment) %>%
  summarise(
    Rhizobacter_Abundance = sum(Abundance),
    Rhizobacter_Relative_Abundance = sum(Relative_Abundance),
    .groups = "keep"
  )

# Agrobacterium
agrobacterium.abundances <- bacteria.abundance.data %>%
  filter(Genus == "Agrobacterium") %>%
  select(Population, Compartment, Abundance, Relative_Abundance) %>%
  group_by(Population, Compartment) %>%
  summarise(
    Agrobacterium_Abundance = sum(Abundance),
    Agrobacterium_Relative_Abundance = sum(Relative_Abundance),
    .groups = "keep"
  )

## Combine abundances into a single tibble with urbanization metrics
pathogenic.bacteria.data <- sample.data %>%
    full_join(mycobacterium.abundances, by = c("Population", "Compartment")) %>%
    full_join(legionella.abundances, by = c("Population", "Compartment")) %>%
    full_join(staphylococcus.abundances, by = c("Population", "Compartment")) %>%
    full_join(stenotrophomonas.abundances, by = c("Population", "Compartment")) %>%
    full_join(acinetobacter.abundances, by = c("Population", "Compartment")) %>%
    full_join(clostridium.abundances, by = c("Population", "Compartment")) %>%
    full_join(paenibacillus.abundances, by = c("Population", "Compartment")) %>%
    full_join(williamsia.abundances, by = c("Population", "Compartment")) %>%
    full_join(sphingomonas.abundances, by = c("Population", "Compartment")) %>%
    full_join(xanthomonas.abundances, by = c("Population", "Compartment")) %>%
    full_join(rhizobacter.abundances, by = c("Population", "Compartment")) %>%
    full_join(agrobacterium.abundances, by = c("Population", "Compartment")) %>%
    replace(is.na(.), 0) %>%
    full_join(urbanization.data, by = "Population") %>%
    type_convert(col_types = c("ffnnnnnnnnnnnnnnnnnnnnnnnnnnn"))
```

Mycobacterium GAMs

```r
## Mycobacterium by distance
mycobacterium.by.distance.GAM <- gam(
    Mycobacterium_Relative_Abundance ~ Compartment
      + s(Distance, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)


## Mycobacterium by HII
mycobacterium.by.HII.GAM <- gam(
    Mycobacterium_Relative_Abundance ~ Compartment
      + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)


## Mycobacterium by ISC
mycobacterium.by.ISC.GAM <- gam(
    Mycobacterium_Relative_Abundance ~ Compartment
      + s(Mean_ISC, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)
```

Check Model Assumptions

```r
## Mycobacterium abundance-by-distance model diagnostics
check_model(mycobacterium.by.distance.GAM)
# Visual check = assumptions met

## Mycobacterium abundance-by-HII model diagnostics
check_model(mycobacterium.by.HII.GAM)
# Visual check = assumptions met

## Mycobacterium abundance-by-ISC model diagnostics
check_model(mycobacterium.by.ISC.GAM)
# Visual check = assumptions met
```

ANOVAs

Table 1: ANOVA table for the Mycobacterium relative abundance-by-distance GAM. Adjusted R-squared = 0.592, deviance = 60.1 Compartment: F = 100, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	0.000	9	0.000	0.644
s(Distance):CompartmentSoil	0.526	9	0.122	0.153

Table 2: ANOVA table for the Mycobacterium relative abundance-by-HII GAM. Adjusted R-squared = 0.589, deviance = 59.8 Compartment: F = 99.4, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0.345	9	0.058	0.221
s(Human_Influence_Index):CompartmentSoil	0.105	9	0.013	0.295

Table 3: ANOVA table for the Mycobacterium relative abundance-by-ISC GAM. Adjusted R-squared = 0.586, deviance = 59.2 Compartment: F = 98.5, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0.000	9	0.000	0.551
s(Mean_ISC):CompartmentSoil	0.067	9	0.008	0.304

Legionella GAMs

```r
## Legionella by distance
legionella.by.distance.GAM <- gam(
    Legionella_Relative_Abundance ~ Compartment
      + s(Distance, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)


## Legionella by HII
legionella.by.HII.GAM <- gam(
    Legionella_Relative_Abundance ~ Compartment
      + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)


## Legionella by ISC
legionella.by.ISC.GAM <- gam(
    Legionella_Relative_Abundance ~ Compartment
      + s(Mean_ISC, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)
```

Check Model Assumptions

```r
## Legionella abundance-by-distance model diagnostics
check_model(legionella.by.distance.GAM)
# Visual check = assumptions met

## Legionella abundance-by-HII model diagnostics
check_model(legionella.by.HII.GAM)
# Visual check = assumptions met

## Legionella abundance-by-ISC model diagnostics
check_model(legionella.by.ISC.GAM)
# Visual check = assumptions met
```

ANOVAs

Table 4: ANOVA table for the Legionella relative abundance-by-distance GAM. Adjusted R-squared = 0.481, deviance = 49.4 Compartment: F = 62.6, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	0.604	9	0.166	0.119
s(Distance):CompartmentSoil	0.427	9	0.082	0.192

Table 5: ANOVA table for the Legionella relative abundance-by-HII GAM. Adjusted R-squared = 0.479, deviance = 49.2 Compartment: F = 62.4, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0.313	9	0.051	0.232
s(Human_Influence_Index):CompartmentSoil	0.000	9	0.000	0.587

Table 6: ANOVA table for the Legionella relative abundance-by-ISC GAM. Adjusted R-squared = 0.478, deviance = 49.1 Compartment: F = 62.3, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0.556	9	0.138	0.139
s(Mean_ISC):CompartmentSoil	0.631	9	0.182	0.111

Staphylococcus GAMs

```r
## Staphylococcus by distance
staphylococcus.by.distance.GAM <- gam(
    Staphylococcus_Relative_Abundance ~ Compartment
      + s(Distance, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)

## Staphylococcus by HII
staphylococcus.by.HII.GAM <- gam(
    Staphylococcus_Relative_Abundance ~ Compartment
      + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)

## Staphylococcus by ISC
staphylococcus.by.ISC.GAM <- gam(
    Staphylococcus_Relative_Abundance ~ Compartment
      + s(Mean_ISC, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)
```

Check Model Assumptions

```r
## Staphylococcus abundance-by-distance model diagnostics
check_model(staphylococcus.by.distance.GAM)
# Visual check = assumptions met

## Staphylococcus abundance-by-HII model diagnostics
check_model(staphylococcus.by.HII.GAM)
# Visual check = assumptions met

## Staphylococcus abundance-by-ISC model diagnostics
check_model(staphylococcus.by.ISC.GAM)
# Visual check = assumptions met
```

ANOVAs

Table 7: ANOVA table for the Staphylococcus relative abundance-by-distance GAM. Adjusted R-squared = 0.117, deviance = 15.4 Compartment: F = 1.836, P = 0.180.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	1.91	9	0.924	0.012
s(Distance):CompartmentSoil	0.00	9	0.000	0.956

Table 8: ANOVA table for the Staphylococcus relative abundance-by-HII GAM. Adjusted R-squared = 0.403, deviance = 44.2 Compartment: F = 2.717, P = 0.104.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	3.523	9	4.989	0.000
s(Human_Influence_Index):CompartmentSoil	0.000	9	0.000	0.921

Table 9: ANOVA table for the Staphylococcus relative abundance-by-ISC GAM. Adjusted R-squared = 0.023, deviance = 5.1 Compartment: F = 1.67, P = 0.201.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0.586	9	0.154	0.128
s(Mean_ISC):CompartmentSoil	0.000	9	0.000	0.822

Stenotrophomonas GAMs

```r
## Stenotrophomonas by distance
stenotrophomonas.by.distance.GAM <- gam(
    Stenotrophomonas_Relative_Abundance ~ Compartment
      + s(Distance, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)


## Stenotrophomonas by HII
stenotrophomonas.by.HII.GAM <- gam(
    Stenotrophomonas_Relative_Abundance ~ Compartment
      + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)


## Stenotrophomonas by ISC
stenotrophomonas.by.ISC.GAM <- gam(
    Stenotrophomonas_Relative_Abundance ~ Compartment
      + s(Mean_ISC, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)
```

Check Model Assumptions

```r
## Stenotrophomonas abundance-by-distance model diagnostics
check_model(stenotrophomonas.by.distance.GAM)
# Visual check = assumptions met

## Stenotrophomonas abundance-by-HII model diagnostics
check_model(stenotrophomonas.by.HII.GAM)
# Visual check = assumptions met

## Stenotrophomonas abundance-by-ISC model diagnostics
check_model(stenotrophomonas.by.ISC.GAM)
# Visual check = assumptions met
```

ANOVAs

Table 10: ANOVA table for the Stenotrophomonas relative abundance-by-distance GAM. Adjusted R-squared = 0.257, deviance = 27.6 Compartment: F = 21.2, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	0.799	9	0.404	0.035
s(Distance):CompartmentSoil	0.000	9	0.000	1.000

Table 11: ANOVA table for the Stenotrophomonas relative abundance-by-HII GAM. Adjusted R-squared = 0.225, deviance = 24.1 Compartment: F = 20.3, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0.406	9	0.076	0.199
s(Human_Influence_Index):CompartmentSoil	0.000	9	0.000	0.862

Table 12: ANOVA table for the Stenotrophomonas relative abundance-by-ISC GAM. Adjusted R-squared = 0.242, deviance = 26.1 Compartment: F = 20.8, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0.697	9	0.248	0.077
s(Mean_ISC):CompartmentSoil	0.000	9	0.000	0.994

Acinetobacter GAMs

```r
## Acinetobacter by distance
acinetobacter.by.distance.GAM <- gam(
    Acinetobacter_Relative_Abundance ~ Compartment
      + s(Distance, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)

## Acinetobacter by HII
acinetobacter.by.HII.GAM <- gam(
    Acinetobacter_Relative_Abundance ~ Compartment
      + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)

## Acinetobacter by ISC
acinetobacter.by.ISC.GAM <- gam(
    Acinetobacter_Relative_Abundance ~ Compartment
      + s(Mean_ISC, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)
```

Check Model Assumptions

```r
## Acinetobacter abundance-by-distance model diagnostics
check_model(acinetobacter.by.distance.GAM)
# Visual check = assumptions met

## Acinetobacter abundance-by-HII model diagnostics
check_model(acinetobacter.by.HII.GAM)
# Visual check = assumptions met

## Acinetobacter abundance-by-ISC model diagnostics
check_model(acinetobacter.by.ISC.GAM)
# Visual check = assumptions met
```

ANOVAs

Table 13: ANOVA table for the Acinetobacter relative abundance-by-distance GAM. Adjusted R-squared = 0.089, deviance = 11.4 Compartment: F = 2.542, P = 0.116.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	0.864	9	0.581	0.015
s(Distance):CompartmentSoil	0.001	9	0.000	0.376

Table 14: ANOVA table for the Acinetobacter relative abundance-by-HII GAM. Adjusted R-squared = 0.028, deviance = 4.8 Compartment: F = 2.381, P = 0.127.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0.379	9	0.068	0.209
s(Human_Influence_Index):CompartmentSoil	0.001	9	0.000	0.385

Table 15: ANOVA table for the Acinetobacter relative abundance-by-ISC GAM. Adjusted R-squared = 0.065, deviance = 8.9 Compartment: F = 2.476, P = 0.120.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0.779	9	0.369	0.042
s(Mean_ISC):CompartmentSoil	0.000	9	0.000	0.655

Clostridium GAMs

```r
## Clostridium by distance
clostridium.by.distance.GAM <- gam(
    Clostridium_Relative_Abundance ~ Compartment
      + s(Distance, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)


## Clostridium by HII
clostridium.by.HII.GAM <- gam(
    Clostridium_Relative_Abundance ~ Compartment
      + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)


## Clostridium by ISC
clostridium.by.ISC.GAM <- gam(
    Clostridium_Relative_Abundance ~ Compartment
      + s(Mean_ISC, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)
```

Check Model Assumptions

```r
## Clostridium abundance-by-distance model diagnostics
check_model(clostridium.by.distance.GAM)
# Visual check = assumptions met

## Clostridium abundance-by-HII model diagnostics
check_model(clostridium.by.HII.GAM)
# Visual check = assumptions met

## Clostridium abundance-by-ISC model diagnostics
check_model(clostridium.by.ISC.GAM)
# Visual check = assumptions met
```

ANOVAs

Table 16: ANOVA table for the Clostridium relative abundance-by-distance GAM. Adjusted R-squared = 0.520, deviance = 53.4 Compartment: F = 68.9, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	0.000	9	0.000	0.793
s(Distance):CompartmentSoil	0.902	9	0.771	0.006

Table 17: ANOVA table for the Clostridium relative abundance-by-HII GAM. Adjusted R-squared = 0.479, deviance = 49.0 Compartment: F = 63.4, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0.000	9	0.000	0.786
s(Human_Influence_Index):CompartmentSoil	0.492	9	0.107	0.165

Table 18: ANOVA table for the Clostridium relative abundance-by-ISC GAM. Adjusted R-squared = 0.506, deviance = 51.9 Compartment: F = 66.85, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0.000	9	0.000	0.655
s(Mean_ISC):CompartmentSoil	0.843	9	0.522	0.020

Paenibacillus GAMs

```r
## Paenibacillus by distance
paenibacillus.by.distance.GAM <- gam(
    Paenibacillus_Relative_Abundance ~ Compartment
      + s(Distance, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)

## Paenibacillus by HII
paenibacillus.by.HII.GAM <- gam(
    Paenibacillus_Relative_Abundance ~ Compartment
      + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)

## Paenibacillus by ISC
paenibacillus.by.ISC.GAM <- gam(
    Paenibacillus_Relative_Abundance ~ Compartment
      + s(Mean_ISC, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)
```

Check Model Assumptions

```r
## Paenibacillus abundance-by-distance model diagnostics
check_model(paenibacillus.by.distance.GAM)
# Visual check = assumptions met

## Paenibacillus abundance-by-HII model diagnostics
check_model(paenibacillus.by.HII.GAM)
# Visual check = assumptions met

## Paenibacillus abundance-by-ISC model diagnostics
check_model(paenibacillus.by.ISC.GAM)
# Visual check = assumptions met
```

ANOVAs

Table 19: ANOVA table for the Paenibacillus relative abundance-by-distance GAM. Adjusted R-squared = 0.396, deviance = 41.3 Compartment: F = 39.78, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	0.000	9	0.000	0.614
s(Distance):CompartmentSoil	0.892	9	0.726	0.008

Table 20: ANOVA table for the Paenibacillus relative abundance-by-HII GAM. Adjusted R-squared = 0.388, deviance = 40.9 Compartment: F = 39.21, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0.617	9	0.179	0.111
s(Human_Influence_Index):CompartmentSoil	0.794	9	0.428	0.031

Table 21: ANOVA table for the Paenibacillus relative abundance-by-ISC GAM. Adjusted R-squared = 0.349, deviance = 36.3 Compartment: F = 36.87, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0.000	9	0.000	0.619
s(Mean_ISC):CompartmentSoil	0.526	9	0.121	0.154

Williamsia GAMs

```r
## Williamsia by distance
williamsia.by.distance.GAM <- gam(
    Williamsia_Relative_Abundance ~ Compartment
      + s(Distance, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)

## Williamsia by HII
williamsia.by.HII.GAM <- gam(
    Williamsia_Relative_Abundance ~ Compartment
      + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)

## Williamsia by ISC
williamsia.by.ISC.GAM <- gam(
    Williamsia_Relative_Abundance ~ Compartment
      + s(Mean_ISC, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)
```

Check Model Assumptions

```r
## Williamsia abundance-by-distance model diagnostics
check_model(williamsia.by.distance.GAM)
# Visual check = assumptions met

## Williamsia abundance-by-HII model diagnostics
check_model(williamsia.by.HII.GAM)
# Visual check = assumptions met

## Williamsia abundance-by-ISC model diagnostics
check_model(williamsia.by.ISC.GAM)
# Visual check = assumptions met
```

ANOVAs

Table 22: ANOVA table for the Williamsia relative abundance-by-distance GAM. Adjusted R-squared = 0.038, deviance = 6.24 Compartment: F = 0.914, P = 0.342.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	0.000	9	0.000	0.772
s(Distance):CompartmentSoil	0.749	9	0.312	0.055

Table 23: ANOVA table for the Williamsia relative abundance-by-HII GAM. Adjusted R-squared = 0.044, deviance = 6.83 Compartment: F = 0.920, P = 0.341.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0.000	9	0.00	0.989
s(Human_Influence_Index):CompartmentSoil	0.764	9	0.36	0.043

Table 24: ANOVA table for the Williamsia relative abundance-by-ISC GAM. Adjusted R-squared = 0.023, deviance = 4.64 Compartment: F = 0.901, P = 0.346.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0.000	9	0.000	0.797
s(Mean_ISC):CompartmentSoil	0.639	9	0.193	0.103

Sphingomonas GAMs

```r
## Sphingomonas by distance
sphingomonas.by.distance.GAM <- gam(
    Sphingomonas_Relative_Abundance ~ Compartment
      + s(Distance, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)


## Sphingomonas by HII
sphingomonas.by.HII.GAM <- gam(
    Sphingomonas_Relative_Abundance ~ Compartment
      + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)


## Sphingomonas by ISC
sphingomonas.by.ISC.GAM <- gam(
    Sphingomonas_Relative_Abundance ~ Compartment
      + s(Mean_ISC, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)
```

Check Model Assumptions

```r
## Sphingomonas abundance-by-distance model diagnostics
check_model(sphingomonas.by.distance.GAM)
# Visual check = assumptions met

## Sphingomonas abundance-by-HII model diagnostics
check_model(sphingomonas.by.HII.GAM)
# Visual check = assumptions met

## Sphingomonas abundance-by-ISC model diagnostics
check_model(sphingomonas.by.ISC.GAM)
# Visual check = assumptions met
```

ANOVAs

Table 25: ANOVA table for the Sphingomonas relative abundance-by-distance GAM. Adjusted R-squared = 0.584, deviance = 60.4 Compartment: F = 76.42, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	2.433	9	2.365	0.000
s(Distance):CompartmentSoil	0.000	9	0.000	0.405

Table 26: ANOVA table for the Sphingomonas relative abundance-by-HII GAM. Adjusted R-squared = 0.453, deviance = 46.1 Compartment: F = 58.2, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0	9	0	0.885
s(Human_Influence_Index):CompartmentSoil	0	9	0	0.998

Table 27: ANOVA table for the Sphingomonas relative abundance-by-ISC GAM. Adjusted R-squared = 0.453, deviance = 46.1 Compartment: F = 58.2, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0	9	0	0.751
s(Mean_ISC):CompartmentSoil	0	9	0	0.774

Xanthomonas GAMs

```r
## Xanthomonas by distance
xanthomonas.by.distance.GAM <- gam(
    Xanthomonas_Relative_Abundance ~ Compartment
      + s(Distance, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)


## Xanthomonas by HII
xanthomonas.by.HII.GAM <- gam(
    Xanthomonas_Relative_Abundance ~ Compartment
      + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)


## Xanthomonas by ISC
xanthomonas.by.ISC.GAM <- gam(
    Xanthomonas_Relative_Abundance ~ Compartment
      + s(Mean_ISC, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)
```

Check Model Assumptions

```r
## Xanthomonas abundance-by-distance model diagnostics
check_model(xanthomonas.by.distance.GAM)
# Visual check = assumptions met

## Xanthomonas abundance-by-HII model diagnostics
check_model(xanthomonas.by.HII.GAM)
# Visual check = assumptions met

## Xanthomonas abundance-by-ISC model diagnostics
check_model(xanthomonas.by.ISC.GAM)
# Visual check = assumptions met
```

ANOVAs

Table 28: ANOVA table for the Xanthomonas relative abundance-by-distance GAM. Adjusted R-squared = 0.251, deviance = 27.1 Compartment: F = 18.21, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	0.886	9	0.656	0.011
s(Distance):CompartmentSoil	0.000	9	0.000	0.922

Table 29: ANOVA table for the Xanthomonas relative abundance-by-HII GAM. Adjusted R-squared = 0.306, deviance = 32.5 Compartment: F = 19.64, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0.922	9	1.302	0.001
s(Human_Influence_Index):CompartmentSoil	0.000	9	0.000	0.794

Table 30: ANOVA table for the Xanthomonas relative abundance-by-ISC GAM. Adjusted R-squared = 0.234, deviance = 25.5 Compartment: F = 17.82, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0.834	9	0.48	0.025
s(Mean_ISC):CompartmentSoil	0.000	9	0.00	0.997

Rhizobacter GAMs

```r
## Rhizobacter by distance
rhizobacter.by.distance.GAM <- gam(
    Rhizobacter_Relative_Abundance ~ Compartment
      + s(Distance, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)


## Rhizobacter by HII
rhizobacter.by.HII.GAM <- gam(
    Rhizobacter_Relative_Abundance ~ Compartment
      + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)


## Rhizobacter by ISC
rhizobacter.by.ISC.GAM <- gam(
    Rhizobacter_Relative_Abundance ~ Compartment
      + s(Mean_ISC, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)
```

Check Model Assumptions

```r
## Rhizobacter abundance-by-distance model diagnostics
check_model(rhizobacter.by.distance.GAM)
# Visual check = assumptions met

## Rhizobacter abundance-by-HII model diagnostics
check_model(rhizobacter.by.HII.GAM)
# Visual check = assumptions met

## Rhizobacter abundance-by-ISC model diagnostics
check_model(rhizobacter.by.ISC.GAM)
# Visual check = assumptions met
```

ANOVAs

Table 31: ANOVA table for the Rhizobacter relative abundance-by-distance GAM. Adjusted R-squared = 0.653, deviance = 67.6 Compartment: F = 115.9, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	3.462	9	1.684	0.003
s(Distance):CompartmentSoil	0.000	9	0.000	0.728

Table 32: ANOVA table for the Rhizobacter relative abundance-by-HII GAM. Adjusted R-squared = 0.589, deviance = 59.9 Compartment: F = 97.81, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0.688	9	0.245	0.078
s(Human_Influence_Index):CompartmentSoil	0.000	9	0.000	0.786

Table 33: ANOVA table for the Rhizobacter relative abundance-by-ISC GAM. Adjusted R-squared = 0.587, deviance = 59.7 Compartment: F = 97.38, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0.66	9	0.21	0.094
s(Mean_ISC):CompartmentSoil	0.00	9	0.00	0.793

Agrobacterium GAMs

```r
## Agrobacterium by distance
agrobacterium.by.distance.GAM <- gam(
    Agrobacterium_Relative_Abundance ~ Compartment
      + s(Distance, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)


## Agrobacterium by HII
agrobacterium.by.HII.GAM <- gam(
    Agrobacterium_Relative_Abundance ~ Compartment
      + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)


## Agrobacterium by ISC
agrobacterium.by.ISC.GAM <- gam(
    Agrobacterium_Relative_Abundance ~ Compartment
      + s(Mean_ISC, by = Compartment, bs = "ts", k = 10),
    data = pathogenic.bacteria.data,
    method = "REML"
)
```

Check Model Assumptions

```r
## Agrobacterium abundance-by-distance model diagnostics
check_model(agrobacterium.by.distance.GAM)
# Visual check = assumptions met

## Agrobacterium abundance-by-HII model diagnostics
check_model(agrobacterium.by.HII.GAM)
# Visual check = assumptions met

## Agrobacterium abundance-by-ISC model diagnostics
check_model(agrobacterium.by.ISC.GAM)
# Visual check = assumptions met
```

ANOVAs

Table 34: ANOVA table for the Agrobacterium relative abundance-by-distance GAM. Adjusted R-squared = 0.539, deviance = 56.4 Compartment: F = 59.84, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	2.65	9	2.438	0.000
s(Distance):CompartmentSoil	0.00	9	0.000	0.934

Table 35: ANOVA table for the Agrobacterium relative abundance-by-HII GAM. Adjusted R-squared = 0.692, deviance = 71.8 Compartment: F = 89.55, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	4.81	9	7.399	0.000
s(Human_Influence_Index):CompartmentSoil	0.00	9	0.000	0.845

Table 36: ANOVA table for the Agrobacterium relative abundance-by-ISC GAM. Adjusted R-squared = 0.474, deviance = 49.9 Compartment: F = 52.36, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	2.325	9	1.189	0.006
s(Mean_ISC):CompartmentSoil	0.000	9	0.000	1.000

Export Data

```r
## Pathogenic bacteria
write_rds(
    pathogenic.bacteria.data,
    file = "data/pathogenic_bacteria_data.rds"
)
```

```r
save.image("data_analysis/10-bacterial_functional_groups/pathogenic_bacteria-workspace.RData")
```

R Session Information

Table 37: Packages required for data management and analysis.

Package	Loaded Version	Date
bayestestR	0.13.2	2024-02-12
broom	1.0.5	2023-06-09
correlation	0.8.4	2023-04-06
datawizard	0.10.0	2024-03-26
dplyr	1.1.4	2023-11-17
easystats	0.7.1	2024-03-28
effectsize	0.8.7	2024-04-01
forcats	1.0.0	2023-01-29
ggplot2	3.5.1	2024-04-23
insight	0.19.10	2024-03-22
kableExtra	1.4.0	2024-01-24
knitr	1.46	2024-04-06
lubridate	1.9.3	2023-09-27
mgcv	1.9-1	2023-12-21
modelbased	0.8.7	2024-02-15
nlme	3.1-164	2023-11-27
parameters	0.21.6	2024-03-18
performance	0.11.0	2024-03-22
purrr	1.0.2	2023-08-10
readr	2.1.5	2024-01-10
report	0.5.8	2023-12-07
see	0.8.4	2024-04-29
stringr	1.5.1	2023-11-14
tibble	3.2.1	2023-03-20
tidyamplicons	0.2.2	2022-09-10
tidyr	1.3.1	2024-01-24
tidyverse	2.0.0	2023-02-22